ROMS Data Assimilation Tools and Techniques

Emanuele Di Lorenzo
School of Earth and Atmospheric Sciences
Georgia Institute of Technology
311 Ferst Drive, Atlanta, GA, 30332-0340
phone: (404) 894-1749 fax: (404) 894-5638 email: edl@eas.gatech.edu

Grant Number: N00014-05-1-0365

LONG-TERM GOALS

Our long-term technical goal is to develop and test the Tangent Linear Model (TLM), Representer Model (RPM), and Adjoint Model (ADM) versions of ROMS (Regional Ocean Modeling System) and TOMS (Terrain-following Ocean Modeling System) for variational data assimilation, ensemble forecasting, and stability analysis. The primary focus is to develop a general platform for strong and weak constraint 4D Variational data assimilation (4DVar); to develop an ensemble prediction capability based on optimal perturbations and stochastic optimals; and to develop stability analysis tools based on eigenmodes and singular vectors to explore the role of environmental stochastic forcing in shaping ocean circulation. Our long-term scientific goal is to model and predict the mesoscale circulation and the ecosystem response to physical forcing in the various regions of the World Ocean through state estimation.

OBJECTIVES

The objectives and scientific goals of the proposed research are:

1. To explore the factors (*e.g*. uncertainties in initial conditions versus those in surface forcing and boundary conditions) that limit the predictability of the circulation in regional ocean models in a variety of dynamical regimes;

2. To develop state-of-the-art variational data assimilation platforms (strong and weak constraint 4DVar) and gain experience in regional ocean applications;

3. To develop ensemble prediction techniques for regional ocean models.

APPROACH

This is a collaborative effort involving Dr. Andrew M. Moore at University of California at Santa Cruz, Drs. Arthur J. Miller and Bruce D. Cornuelle at the Scripps Institution of Oceanography, and Dr. Hernan Arango at Rutgers University. To address the aforementioned goals and objectives, we are using a newly developed suite of tools that utilize ROMS/TOMS tangent linear and adjoint models. These models and tools were developed under the support of previous ONR funding.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **30 SEP 2007**	2. REPORT TYPE **Annual**	3. DATES COVERED **00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE **ROMS Data Assimilation Tools And Techniques**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Georgia Institute of Technology,School of Earth and Atmospheric Sciences,311 Ferst Drive,Atlanta,GA,30332**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
code 1 only

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**13**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

To address objective (**1**), we have used the ideas of Generalized Stability Theory (GST) in order to identify the most unstable directions of state-space in which errors and uncertainties are likely to grow. Specifically, for initial and boundary condition errors we compute the singular vectors of the TLM propagator, while for surface forcing we compute the stochastic optimals. By examining the details and dynamics of these structures we can learn much about the factors that limit the predictability of the circulation.

To address objective (**2**), we have used several 4-dimensional variational data assimilation schemes that have been developed for ROMS/TOMS. For cases in which the dynamics are imposed as a strong constraint (*i.e.* no model error assumed) we use an incremental 4DVar approach (IS4DVAR) similar to that used operationally at some numerical weather prediction centers. In the case where errors are admitted in the model we use an indirect representer-based weak constraint 4DVar algorithm (W4DVAR) and a weak constraint Physical Space Analysis System (W4DPSAS). W4DVAR is based on the Oregon State University Inverse Ocean Model (IOM) of which ROMS/TOMS is also a component (Di Lorenzo et al, 2007; Muccino et al, 2007). The IOM requires an additional version of the model that computes a finite amplitude linear estimate of the total state of the system as opposed to perturbations about some existing solution of the nonlinear ROMS. This second linearized form of ROMS/TOMS (denoted here as RPM) has been developed under a separate NSF funded effort.

To address objective (**3**), we have used the optimal structures identified in (**1**) using GST to construct ensembles of model forecasts following the approach used operationally at some numerical weather prediction centers.

WORK COMPLETED

Since the start of this current award the following tasks have been completed:

a. Rewritten the hand-coded TLM, RPM, and ADM algorithms of ROMS/TOMS from F77 to F90/F95 to facilitate multiple levels of nesting, parallelization and improve computational efficiency. The TLM, RPM and ADM are approximately 2.5, 2.6 and 2.8 times more computationally expensive than the Nonlinear Model (NLM). These algorithms were updated to the latest version of ROMS/TOMS framework, tested and distributed to selected beta-testers around the world on May 15, 2006.

b. Parallelized TLM, RPM, and ADM models and their associated drivers. The TLM and RPM can be run in either shared-memory (OpenMP) or distributed-memory (MPI). Currently, the ADM can be only run in distributed-memory because its hand-written construction violates shared-memory mutual exclusion rules between tiles. The TLM and RPM have a parallel structure identical to the NLM. The ADM required adjoint communication exchanges between parallel domain decomposition tiles.

c. Developed and tested the strong constraint, incremental 4DVar (IS4DVAR) following the approach of Courtier et al. (1994). The modeling of the background error covariance uses the generalized diffusion method proposed by Weaver and Courtier (2001). The free parameters controlling the shape of the Gaussian correlation for each state variable are the horizontal and vertical decorrelation length-scales and the diffusion coefficients. The diffusion operators for each state variable are solved explicitly and implicitly. Since the oceanic vertical decorrelation scales are much smaller than the horizontal, the implicit algorithm is preferable and cheaper. It is unconditionally stable for any vertical convolution time-step. The spatially dependent normalization coefficients used to convert the covariance matrix into a correlation matrix are computed using the exact (expensive) or randomization (cheaper) methods (Weaver and Courtier, 2001). These normalization coefficients ensure the diagonal elements of the background error covariance to be equal to unity. After extensive testing with idealized twin experiments, we are now working with realistic data assimilation experiments and state estimation in our US east and west coast, Gulf of Mexico, Caribbean Sea, and East Australia Current applications. We are assimilating various types of data including altimetry, SST from satellites, CTD, XBT, ADCP, and gliders. An operational forecast website for the Gulf of Mexico has been made available to the public (http://www.myroms.org/applications/ias/) and is described in Powell et al. (2007).

d. Developed the weak constraint 4DVar (W4DVAR) (Di Lorenzo et al., 2007) single and multiple drivers using the indirect representer approach described in Chua and Bennett (2001). The multiple driver option is used to interface with the IOM framework and developed under separate NSF/ITR funding. As in IS4DVAR, the model error covariance is modeled with the generalized diffusion operator. Currently, we are using the W4DVAR drivers in our US west coast and Intra-Americas Sea applications.

e. Developed a weak constraint 4D-PSAS (W4DPSAS) driver. The PSAS algorithm is similar to W4DVAR but with the RPM replaced with the NLM. That is, the representer functions are not explicitly computed. The PSAS acronym is misleading but it is retained for historical reasons. As in W4DVAR, the minimization is in observation space. Courtier (1997) shows the duality between 4DVar and PSAS. Both algorithms produce identical results if the measurement functional is linear and the background and observation error covariance are the same.

f. Developed an ensemble prediction method based on the ROMS adjoint sensitivity model to quantify the predictability of mesoscale coastal flows in the California Current System (Mosca et al., 2007).

RESULTS

A 3D baroclinic coastal upwelling test case reveals potentials for high forecast skill
The representer-based weak and strong constraint 4DVAR ROMS where used in a realistic 3D baroclinic upwelling system with complex bottom topography (Di Lorenzo et al., 2007). In this example, the flow field is nonlinear and characterized by mesoscale activity as evident from the filamentary and cyclonic structures in the circulation (Fig. 1). Synthetic observations of upper ocean (0-450m) temperatures and currents with a high resolution sampling array (referred to as the HIRES sampling) were assimilated over a 10 day window. Both the strong and weak constraint inverse solutions were able to greatly reduce the initial error variance by 97% and 80% respectively (Fig. 2). We also found that both solutions exhibit relatively high forecast skill when used to initialize the nonlinear model at the end of the assimilation window. Significant forecast skill was found up to 10-20 days after the last observation is assimilated and is higher than the persistence timescale of the flow, which for this upwelling regime is less than 5 days (Fig. 2).

The same experiment was repeated using an observing array that is both spatially and temporally aliased (referred to as the COARSE sampling). For this case, both the strong and weak constraint posses similar levels of hindcast and forecast skill, although the weak case was slightly better. During the forecast, the skill was not as high as in the HIRES case, however the spatial pattern correlation with the true is still very high when compared to the first guess (Fig 3). These results suggest that the assimilation platform based on the indirect representer method with inverse ROMS is able to extract the dynamically active information from the observations during the hindcast window and generate a good initialization for the forecast.

The application of the assimilation platform to this highly nonlinear example brought to our attention one important aspect of the methodology that required further consideration. Specifically we found that the use of the linearized ROMS model (RP-ROMS) in an iterative approach (the outer loop in the representer method) does not necessarily converge to the solution obtained by the nonlinear model. This is an assumption that is usually made in the representer method. Instead we found that RP-ROMS is linearly unstable when the flow field is very nonlinear. This implies that in the assimilation window the corrections provided by the inverse solution need also to damp the linear instabilities that develop in the RP-ROMS.

Real-time forecast for the Southern California Current System and Intra-Americas Seas

Applications of inverse ROMS using real ocean observations are currently being performed for the Southern California Bight (Di Lorenzo et al., in prep.) and for the Intra-Americas Seas (Powell et al., 2007). For the California Current System, fig. 4 shows our first real-time forecast for the oceanic conditions off Southern California during May 2006. The forecast was initialized using the ROMS weak constraint assimilation platform in combination with cruise CTD upper ocean data for the month of April 2006. The forecast was posted on the web (http://www.o3d.org/web/CalCOFI/april2006/) on April 30, 2005 and was used to guide the biological sampling during a subsequent cruise in May 2006. Analysis of the forecast skill reveals that the model was able to predict the displacement of the upwelling front (evident in Fig. 4) during mid-may. This work is currently being prepared for publication.

For the Intra-Americas Seas, the strong constrain platform was used to implement a real-time forecasting system that provides daily forecasts accessible through the website http://www.myroms.org/applications/ias/. Fig. 5 shows an image of a forecast for September 28, 2007.

Predictability of mesoscale flows in the coastal ocean

The ROMS adjoint machinery was used to compute optimal perturbation patterns that were used to generate a large ensemble of model forward. This ensemble was than used to characterized the timescales and dynamics of predictability of the mescoale flows in the California Current System (Mosca et al., 2007). IN particular we quantified how errors in the forcing, initial and boundary condition affect the range of predictability. As an example, figure 6 shows that the timescales of predictability for the SSH and SST fields associated with the surface forcing functions have a strong seasonal dependence compared to the predictability range associated with the persistence of initial conditions. The typical persistence timescale is about 15 days for the SST and 20 days for the SSH with little dependence on the season when the forecast is initialized. Errors in the forcing surface functions greatly affect the predictability range during winter/spring, when the forcing is stronger and the upwelling fronts are developing in the California Current. However in late spring and summer when the forcing is weaker, and large-scale eddies are generated, the range of predictability is no longer controlled by the surface forcing but by errors in the initial conditions (Mosca et al., 2007).

IMPACTS/APPLICATIONS

The newly developed ROMS/TOMS adjoint-based platforms are powerful tools for ocean prediction, adaptive sampling, and understanding the underlying circulation dynamics.

TRANSITIONS

The work completed here are now part of the ROMS/TOMS utilities that are freely available to both research and operational communities.

RELATED PROJECTS

The work described here is in collaboration with Dr. Andrew Moore at University California at Santa Cruz, Drs. Arthur Miller and Bruce Cornuelle at the Scripps Institution of Oceanography, and Dr. Hernan Arango at Rutgers University. These investigators are supported by the following grants:

"ROMS/TOMS Tangent Linear and Adjoint Models: Testing and Applications", PI Andrew Moore, grant number N00014-01-1-0209.

"ROMS/TOMS Tangent Linear and Adjoint Models: Testing and Applications", PI Hernan Arango, grant number N00014-05-1-0366.

"ROMS/TOMS Tangent Linear and Adjoint Models: Data Assimilation Tools and Techniques", PIs Arthur Miller and Bruce Cornuelle, grant number N00014-99-1-0045.

Also relevant is a collaborative project involving the PI, Dr. Andrew Moore at University California at Santa Cruz, Dr. Hernan Arango at Rugers University, and Dr. Ralph Milliff at Colorado Research Associates:

"Ocean State Estimation and Prediction of the Intra-Americas Sea", PIs: Andrew Moore, Ralph Milliff and Hernan Arango, grant numbers: N0014-05-M-0277, N00014-05-M-0081, N00014-05-0275.

PUBLICATIONS

Di Lorenzo, E., Moore, A., H. Arango, Chua, B. D. Cornuelle, A. J. Miller, B. Powell and Bennett A., 2007: Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS): development and application for a baroclinic coastal upwelling system. Ocean Modeling, 16 (3-4): 160-187.

Muccino, J., A. Bennett, B. Cornuelle, B. Chua, E. Di Lorenzo, et al., 2007: The Inverse Ocean Modeling System. II: Applications. Journal of Atmospheric and Ocean Technology, in revision.

Moore, A., H. Arango, E. Di Lorenzo, B. D. Cornuelle and A. J. Miller, 2007: An Adjoint Sensitivity Analysis of the Southern California Current Circulation and Ecosystem. Part I: The Physical Circulation. Journal of Physical Oceanography, in revision.

Moore, A., H. Arango, E. Di Lorenzo, B. D. Cornuelle, A. J. Miller, and D. J. Neilson, 2004: A Comprehensive Ocean Prediction and Analysis System Based on the Tangent Linear and Adjoint of a Regional Ocean Model. Ocean Modeling, 7, 227-258.

Mosca, C., E. Di Lorenzo, A. M. Moore, A. J. Miller and H. Arango, 2008: Predictability of mesoscale flows in the California coastal ocean. J. Phys. Oceanogr., in prep.

Powell B. S., H. G. Arango, A. M. Moore, E. Di Lorenzo, R. F. Milliff and D. Foley, 2007: 4DVAR Data Assimilation in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). Ocean Modeling, submitted.

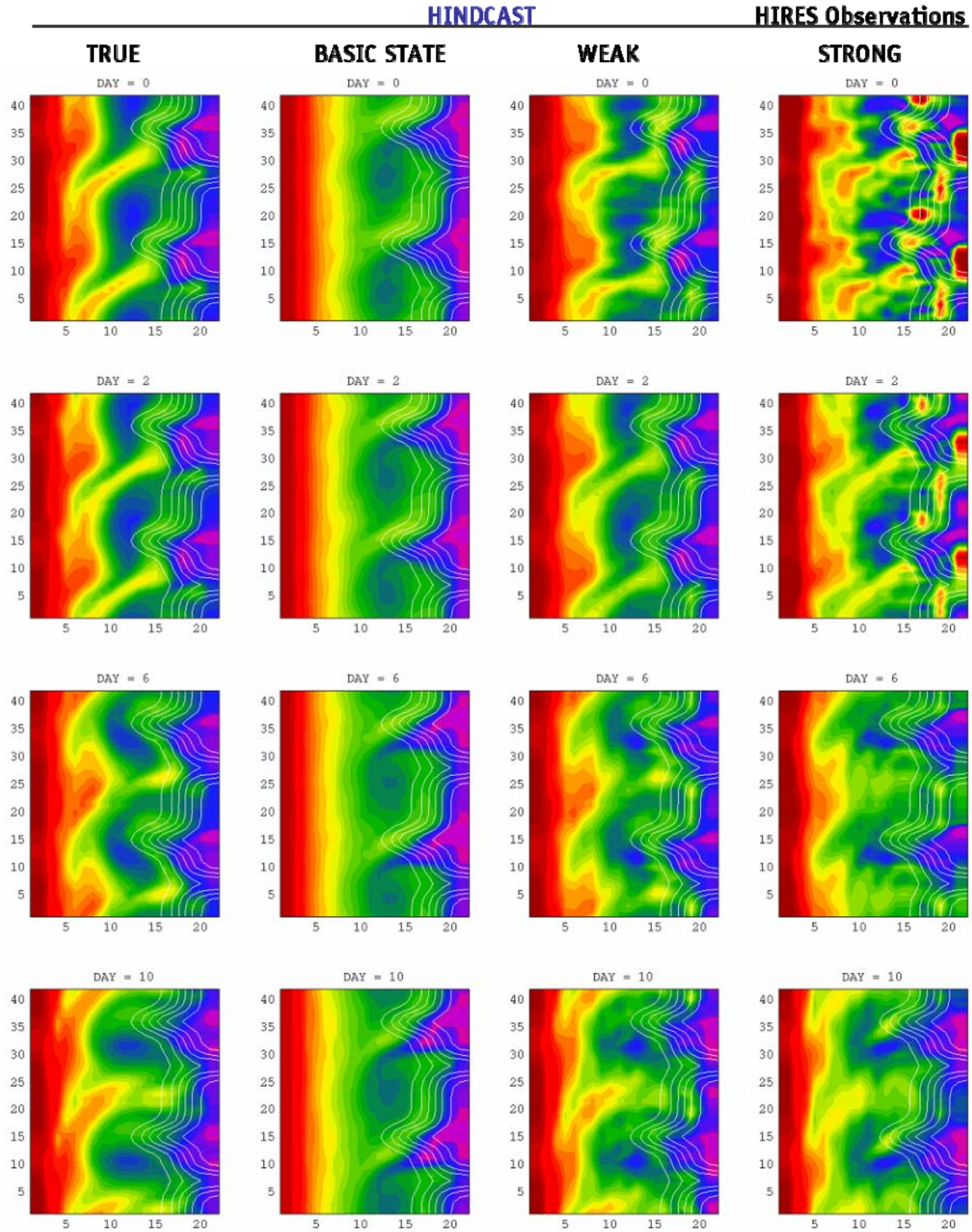

Figure 1: Maps of upper ocean temperature (0-100 m) for hindcast period from DAY=0 (initial condition) to DAY=10 (end of assimilation window). The first column represent the "true" state, the second column is the basic state integration initialized from climatology, the third column is the result from the weak constraint assimilation experiment (Exp_weakH) and the fourth column is the result from the strong constraint assimilation experiment (Exp_strongH)

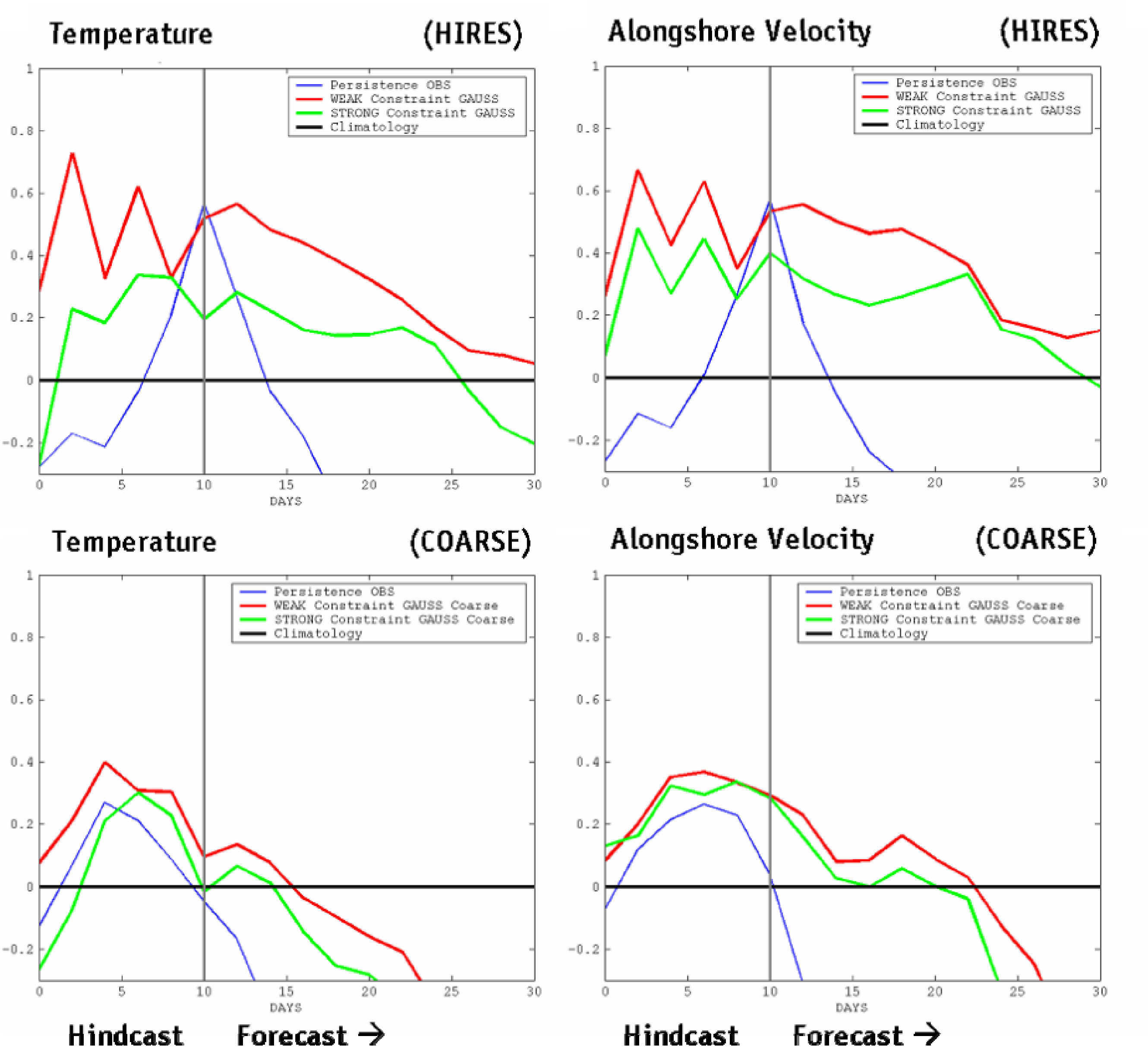

Figure 2: Hindcast and forecast skill scores for upper ocean temperature (0-100 m) and alongshore velocity. Skill is defined based on the RMS difference from the truth (see section 5.6 in paper). A perfect skill value is 1. The red line is the skill of the weak constraint solution and the green line of the strong constraint. The blue line corresponds to the skill of persistence. The field used to compute the persistence skill is computed by taking the available observations and performing an objective mapping to interpolate at the location where observations are not available. In the forecast window (DAY 10-30) no observations are used to constrain the model trajectory. The upper row shows result using the HIRES sampling array and the lower row the COARSE array.

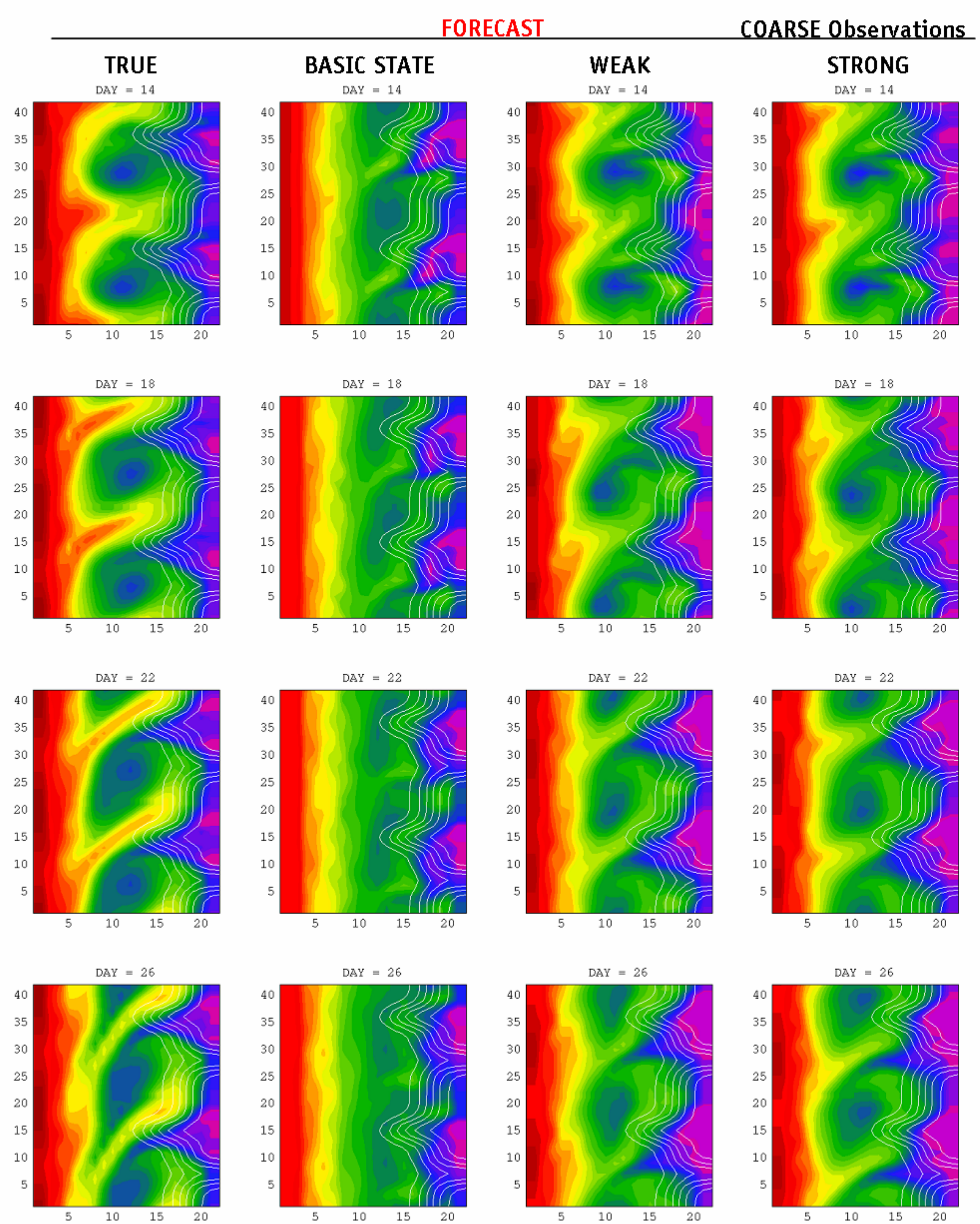

Figure 3: Maps of upper ocean temperature (0-100 m) in the forecast window from DAY=14 (4 days after the end of the assimilation window) to DAY=26. The first column represent the "true" state, the second column is the basic state integration initialized with climatology, the third column is the result from the weak constraint assimilation experiment (Exp_weakC) and the fourth column is the result from the strong constraint assimilation experiment (Exp_strongC). The forecast are performed using the nonlinear model initialized with the solutions of the assimilation experiments at DAY=10.

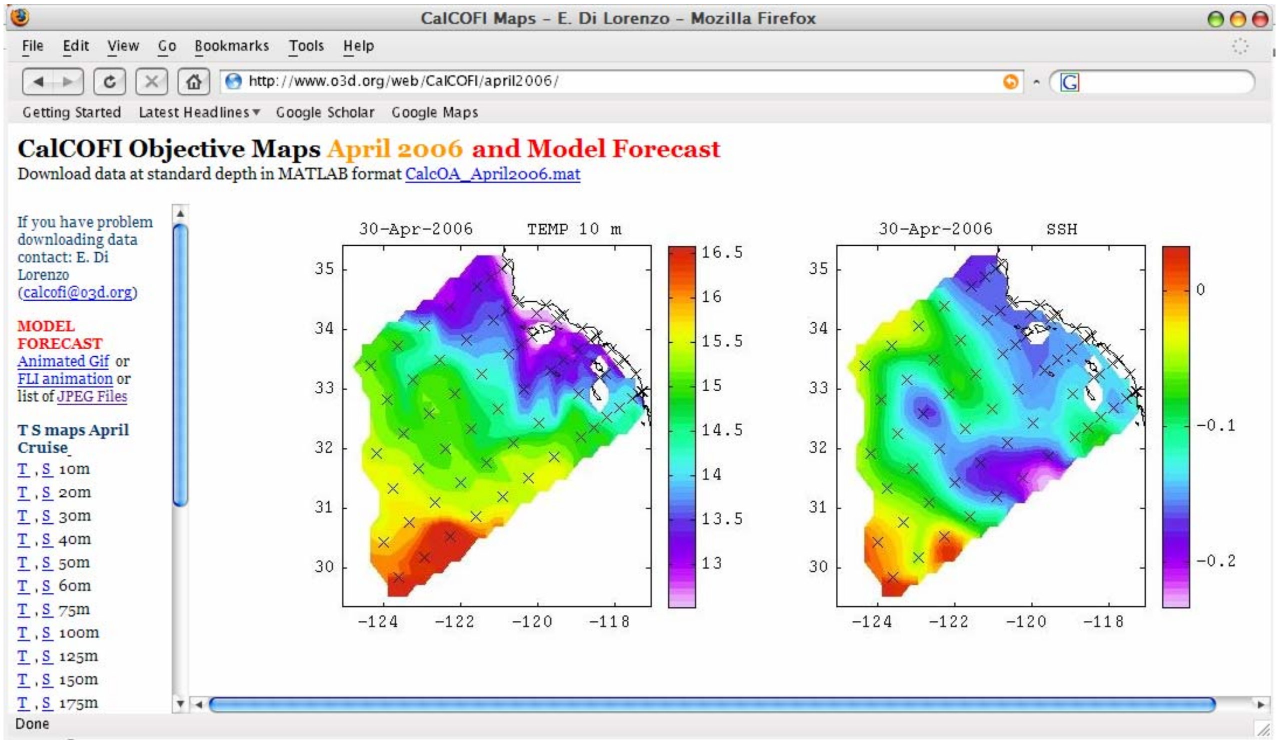

Figure 4: An image of the on-line real time ocean forecast for May 2006. The forecast was initialized using 4D variaitonal data assimilation with the inverse Regional Ocean Modleing System (ROMS). During the assimilation time window, TS data from the April 2006 CalCOFI cruise where assimilated to initialize the model field for the forecast. In the forecast, the model was forced with winds provided by the ECPC regional atmospheric model forecast. This forecast was used to guide biological oceanographers during a cruise in May 2006.

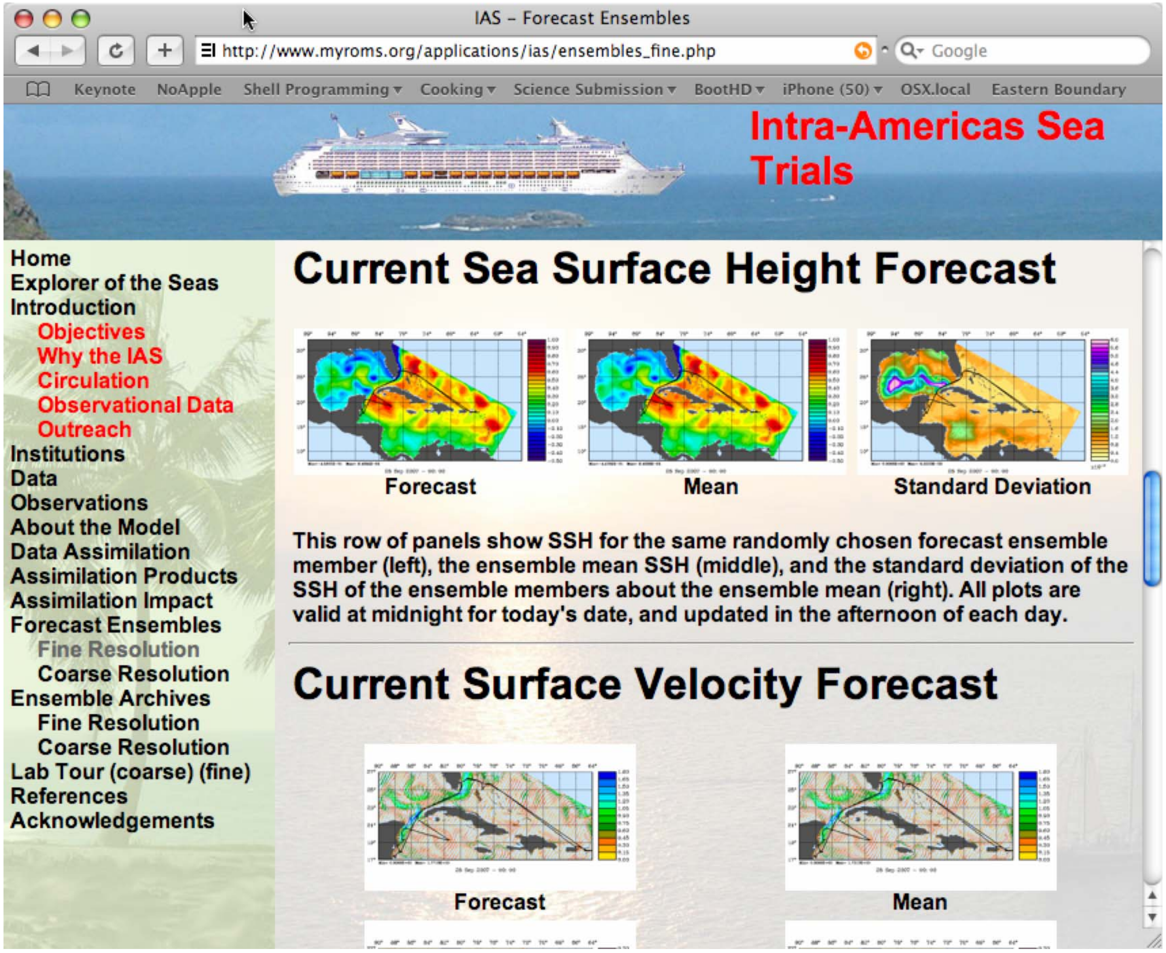

Figure 5: An image of the on-line real Intra-Americas ocean forecasting system (Powell et al., 2007). Plotted are the forecasts for September 28, 2007. The forecast was perfomed using the strong contraint 4D variaitonal data assimilation drivers of Regional Ocean Modleing System (ROMS).

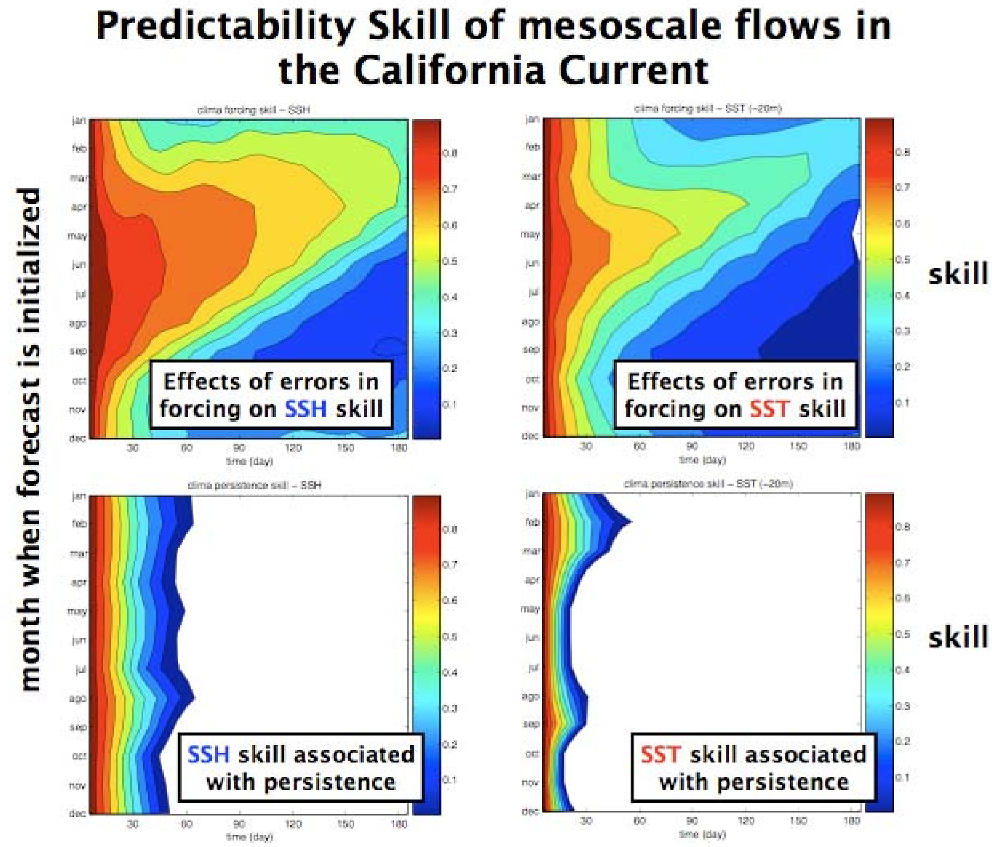

Figure 6: Timescales of predictability for the SSH and SST fields associated with the surface forcing functions and persistence of initial conditions. The typical persistence timescale is about 15 days for the SST and 20 days for the SSH with little dependence on the season when the forecast is initialized. Errors in the forcing surface functions greatly affect the predictability range during winter/spring when the forcing is stronger and the upwelling fronts are developing in the California Current. However in late spring and summer when the forcing is weaker and large-scale eddies are generated the range of predictability is no longer controlled by the surface forcing but by errors in the initial conditions (Mosca et al., 2007).